

DOCUMENT RESUME

ED 275 169

FL 016 073

AUTHOR Klein-Braley, Christine
 TITLE Advance Prediction of Difficulty with C-Tests.
 PUB DATE 84
 NOTE 18p.; In: Practice and Problems in Language Testing. Papers from the International Symposium on Language Testing (7th, Colchester, England, 1984); see FL 016 066.
 PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
 EDRS PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS *Cloze Procedure; Difficulty Level; English (Second Language); Foreign Countries; German; *Item Sampling; *Language Tests; Predictor Variables; Second Languages; *Test Construction; *Testing Problems; *Test Items
 IDENTIFIERS *C Tests; West Germany

ABSTRACT

This report investigates the selection of appropriate texts for C-Tests, a modified form of the cloze test, for assessing second language learning. The procedure for textbook readability first involved the administration of different texts to sample groups to determine the C-test difficulty of individual texts. At the same time, a variety of different indices for each text were calculated. The indices were based on the texts' sentence lengths and the specific problems of the native German-speaking students in understanding the texts. The results were validated in subsequent test administrations to learners of German as a second and as a foreign language and with two groups of German and Finnish learners of English as a second language. It is concluded that this method is useful for predicting the difficulty of a text for use in C-Tests, and that the results imply the existence of absolute text difficulty (C-processing difficulty). This method may also serve as a new statistical model for examining the relationship between C-processing difficulty and the learner's position on the language learning continuum. (MSE)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

FD275169

ADVANCE PREDICTION OF
DIFFICULTY WITH C-TESTS

CHRISTINE KLEIN-BRALEY
(UNIVERSITY OF DUISBURG)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

P. Culhane

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

FL 016 073

ADVANCE PREDICTION OF DIFFICULTY WITH C-TESTS

Christine Klein-Braley (University of Duisburg)

Perhaps the major criticism levelled at my study of cloze tests (Klein-Braley 1981a) by my linguistic assessor was that I had failed to go sufficiently deeply into the question of text selection. This was a justified criticism. However it is not a criticism that applies exclusively to my work. In my search through the relevant literature I was unable to find anywhere guidelines for text selection. Even Oller, who devotes a whole chapter of his *Language tests at school* (1979), to cloze tests fails to provide the necessary help. By implication he suggests indeed that any and every text will serve the purpose. Yet when I began to develop cloze tests myself I found that there were in fact many more unsuitable texts in the world than suitable ones. It seemed to me at the time that there were two aspects to this problem. One was the question of text content, and the other was the question of text difficulty.

On the content side, it seems unfair to use texts which demand specialised knowledge from the examinee, particularly since in the classical cloze test only one long text is involved. However it is difficult to decide where general knowledge stops and specialised knowledge begins. Literary texts seem unsuitable, and older literary texts (Shakespeare?) seem doubly unsuitable, since in these cases an aesthetic and possibly an archaic factor is at work. Humorous and highly colloquial texts also cause problems because of their special linguistic characteristics. What is then left tends to be the contemporary, non-literary, discursive, argumentative or narrative text with no particular idiosyncracies or special terminology. What could, in brief, be called an unmarked text.

So far as text difficulty is concerned, the major assumption seems to be that the teacher or tester developing the test can select 'suitably difficult' texts for his examinees. Again, no additional guidelines are given, although Oller does claim that teacher judgements of difficulty are vastly superior to other more objective techniques, such as readability formulae, for example.

So, in cloze tests, we have a test form for which reliability and validity are claimed to be generic, while at the same time the basis for the tests, namely the texts, are a matter of subjective decisions.

Since the C-Test is a development of the cloze test, or, as we would prefer to formulate it, a different and more satisfactory operationalisation of the construct, we too are confronted with questions of text and test content and with questions of text and test difficulty.

We have not, so far, made any progress in the question of text

content. We do submit the texts selected by one of us to other judges and ask them whether they find the texts sufficiently neutral or unmarked to be used in a C-Test. But other than this we have not been able to define the quality of unmarkedness more exactly. Where we do feel we have made progress is in the area of text difficulty.

The C-Test is a norm-oriented test, and as such, aims at a medium level of difficulty on average. The mean score should be around 50% in order to ensure maximum differentiation between subjects. Inside the C-Test, however, it is preferable to order the texts in ascending difficulty, starting with an icebreaker and finishing with a really tough item. If necessary we can afford to let the mean difficulty slide up to 60%: C-Tests have been shown to have high reliabilities even when they were much too difficult or too easy for the subjects involved and we feel it is important for the icebreaker to be so simple that every test subject understands exactly what the C-Principle demands from him or her.

Since the subjects involved so far in C-Tests have ranged from 7-years-olds to adults, achieving medium difficulty in the tests for every subject group means that we need texts with a variety of difficulty levels, and that subjective judgements of appropriacy are impossible to make. In addition we have been developing tests for three different groups of addressees - L1 learners, L2 learners and LF learners. Moreover we have been working with two different languages - English and German - and this complicates matters rather.

For L1 groups we have been surprisingly successful in constructing appropriate tests, but this is because we cheated. Here we used other people's judgements and chose the texts from readers for the target groups, and combined them with texts from readers aimed at older and younger age groups. Obviously the authors and editors of such text books had some sort of procedure, either objective or subjective, which helped them to determine the adequacy of the texts. The area in which we had major problems was the LF advanced groups in Duisburg - learners of English - and I succeeded in developing appropriately difficult tests for them only by a process of trial and error. However, this is a time-consuming procedure, particularly when one is working inside a system where sufficiently large groups present themselves for testing only once a year. We felt that there had to be a better method than subjective judgements.

My first step was to discover what had been tried out so far. These methods can be grouped into text-oriented and subject-oriented techniques. The subject-oriented techniques include ratings, cloze tests and multiple-choice tests, and they all involve confronting the subject and the text in some way. He must rate the texts, or complete cloze tests constructed from the texts, or answer questions on the texts. Ratings are notoriously problematical (cf. Bartenwerfer & Raatz 1979) and in a small pilot study I found that neither German university students of English nor their academic teachers were particularly unanimous about their judgements of texts with regard to difficulty. Cloze tests were non-starters because both Charles Alderson (1979) and I had shown that the n cloze tests that can be con-

structed from a text using *n*th word deletion are all different tests. In addition I had been able to demonstrate that the difficulty level of a cloze test correlates with the number of structure words deleted. But since the different starting points delete different proportions of structure words, there is no such thing as the 'cloze difficulty' of a text. Bensoussan (1982) was able to show that multiple-choice comprehension questions have a difficulty level of their own, which is not necessarily related to the difficulty level of the text they are attached to, and in an - as yet unpublished - paper presented to the ACROLT Symposium in Israel in 1983 she was able to pin down the technical changes which made the items easier or more difficult.

The text-oriented methods all make use of statistical characteristics of text to predict difficulty levels for specific groups of addressees in advance. In general they take the form of multiple regression equations where two or more indicators are used to predict text difficulty in advance without ever confronting the text and the reader. However there are two major snags. The first is that they produce entirely different predictions. In a large-scale study Stokes (1980) showed that there was no one pair of formulae which consistently predicted the same difficulty levels. In an exploratory study (Klein-Braley 1981b) I was able to show that formulae predictions also failed to agree with ratings or with cloze test results. The second snag is that the validation criteria for the text-oriented methods were in all cases the subject-oriented methods whose shortcomings I have just reported.

One of the problems in readability research is that the construct of readability simply has not been adequately defined. This is shown, for example, by the fact that the readability researchers are not even unanimous about such matters as the percentage of correct answers in a multiple-choice reading test which they will consider as adequate for demonstrating that comprehension has taken place. Some insist on 50% right answers, others on 75%, others on 90%.

Another approach to readability would be to take the findings from text linguistics and apply these to texts. Unfortunately, as Pugh pointed out in 1981, these methods are still very much in their infancy, and so far they do not appear to be sufficiently reliable and replicable for use by the layman. Furthermore the question of text difficulty has not been central to such research.

For this reason the readability research seemed to us to be much more relevant to our problem. These approaches had the advantage of being relatively simple, even though we were not satisfied with the results produced. However our problem, though related, was not identical to that of readability researchers. Readability does not really interest us at all. Our aim is much more modest. All we want to be able to do is to predict in advance how difficult a given text is likely to be for a specific group of addressees. Our criterion is quite simply the mean score obtained by that group on a text mutilated in accordance with the C-Principle.

We therefore decided to use the same approach as that of the readability researchers, i.e. to attempt to relate the statistical characteristics of the texts and the empirically determined C-Test difficulties. A survey of the relevant literature provided a large number of indices which had been used in previous research. We immediately discarded all approaches based on word lists, since these are available only for English, and even here apply only to L1 groups. For German there are no such word lists, and a simple translation of the English lists would not necessarily produce equivalent procedures. We also chose to disregard any indices which demand too much human attention. We had a very small personal computer at our disposal, and it seemed fitting that such things as counting and matching should be done by machine rather than by woman. Since almost all regression equations available include one index which is in some way related to syntactical complexity - for instance sentence length - and one which is related to the range of vocabulary, we ensured that such indices were included in our analysis. The type-token ratio seemed particularly attractive as a vocabulary index since it is not language-bound, can be easily worked out by the computer, and has been advocated by respected scholars working in the field of linguistics (cf. Herdan 1966).

Our procedure was quite simple. We first administered a number of texts to various sample groups and determined the C-Test difficulties for the individual texts. At the same time we calculated a variety of different indices for the texts. We decided to begin with a two predictor regression equation. In order to produce this we next correlated the text indices with the empirically determined C-Test difficulties (P Values). The two indices which had the highest correlations with the C-Test difficulties were then combined in a multiple regression, assuming for the moment that there was no suppressor effect operating.

There are many more Germans and learners of German available for testing in Germany than there are native speakers and learners of other languages. For this reason our German results are much better substantiated and complete than our results for native speakers and learners of other languages. I would therefore like to begin with the German groups.

Last year in Duisburg we were able to present a regression equation for German L1 speakers. The two predictors were the type-token ratio and the number of sentences in the text. This regression equation was derived for third grade pupils on the basis of tests administered to fairly large groups. However we later decided that the number of sentences in a text was very much the result of a subjective decision by us as to where the logical unit 'text' should end. Often one can manipulate these values by simply adding or subtracting a sentence. Since manipulation was not the point of the exercise we decided to use an index which correlated very high with the number of sentences in the text, namely the average length of the sentences in words. An average tends to be less susceptible to minor changes in text length.

Using empirical data for 12 texts and a total of 201 pupils in the third grade we calculated a regression equation for prediction of empirical difficulty in this group using these two indices:

$$Y' = 238 - 203 \text{ TTR} - 1.9 \text{ XSL(W)} \\ (\text{Multiple } R = .76)$$

We have recently conducted experiments in 5th grade classes. (see Raatz, this volume). Once again we took the opportunity to calculate a regression equation. This time it takes the form:

$$Y' = 182 - 142 \text{ TTR} - 0.83 \text{ XLS(W)} \\ (\text{Multiple } R = .69)$$

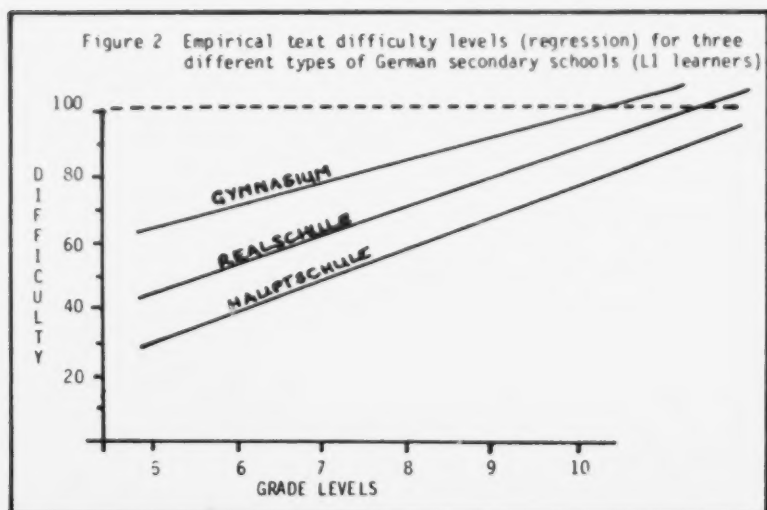
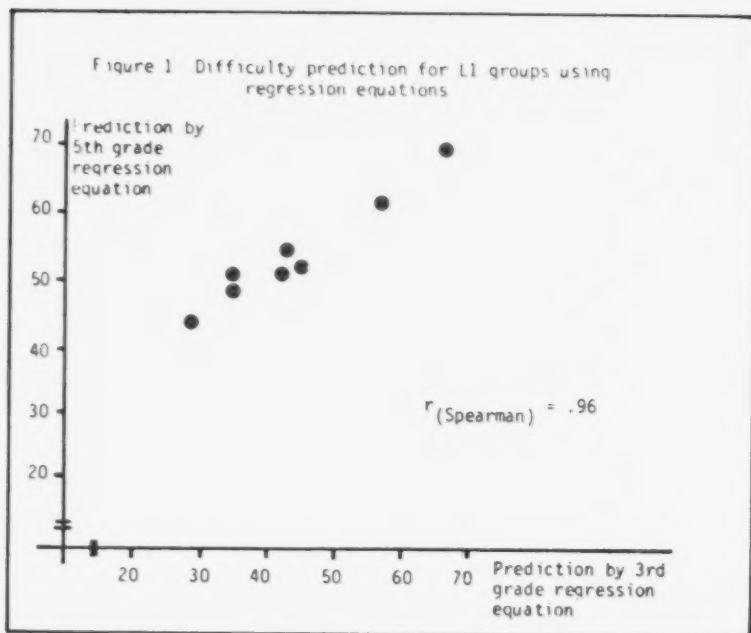
Now these two equations differ, but they have to differ, since in the first case we are predicting 3rd grade difficulty, in the second we are predicting 5th grade difficulty.

The obvious way to check on the accuracy of our predictive formulae would be to administer a number of new texts to 3rd and 5th grade groups. We have not been inactive in the two years since we last presented a report here in Colchester, but there are still only two of us, and our enthusiasm for the C-Test, while enormous, is not unlimited. So we have not been able to do this yet. However we do possess a certain amount of evidence that the formulae work.

The first proof comes from the predictors themselves. Obviously they differ since we are predicting to a different grade level in each case. However both formulae use the same two indicators: the ones which correlate highest with the difficulty indices for the group involved. The relative importance of the indices in the prediction can be ascertained by examining the beta weights. These are virtually identical in both cases. TTR enters with .65 in the third grade and with .62 in the fifth grade. Mean sentence length in words has a beta weight of .39 in the third grade and .41 in the fifth. So not only are the same indicators being used, they also have the same importance in the two equations. Only eight texts were identical in the two investigations, the others were different for the two groups.

We can check the agreement between the two formulae by comparing the prediction of the text difficulty that they produce. Using the eight common texts we calculated difficulty predictions using both formulae and then calculated a rank correlation coefficient for agreement and came up with a value of .96. This is shown in Figure 1 which shows not only the very high agreement between the values, but also the fact that the only real difference between the two values consists of an additive constant, reflecting the fact that the texts become easier as the children grow older. In other words: these equations make identical predictions.

We therefore have three proofs for the accuracy of our predictions: the same two indicators enter into the regression equations as predictors; they have the same beta weights; and if the age increase is taken into consideration they produce the same conclu-



sions since they rank the texts virtually identically in terms of difficulty.

The fourth proof is empirical. For this C-Tests were administered to 19 classes in a Hauptschule, a Realschule and a Gymnasium. In the Hauptschule and the Realschule one class for each grade level between 6 and 10 was selected; in the Gymnasium in addition the 11th grade. In each of the 19 classes empirical difficulties were determined for 12 texts and for each text a prediction was calculated using the regression formula for the 5th grade. Finally the rank correlation between the empirically determined difficulty levels and the prediction produced by the regression equation was calculated.

The results are shown in Table 1.

Table 1

Results for the 12 texts in 19 school classes

School & class	N	Rank correlation between predicted and empirical difficulty values
H 6	17	.74
H 7	21	.54
H 8	19	.71
H 9	29	.72
H 10	20	.67
R 6	29	.83
R 7	23	.74
R 8	26	.70
R 9	27	.69
R 10	30	.83
G 6	26	.62
G 7	31	.57
G 8	34	.59
G 9	30	n.s.
G 10	27	n.s.
G 11	9	n.s.

All the rank correlations for the individual classes are significant with the exception of those for the Gymnasium classes 9, 10 and 11. In all three cases, however, the average total score is around 95%, so that there is no real variation in difficulty between the individual texts. The other correlations range between .54 to .83 with a median value of .70. When the empirical data are pooled the correlation between empirical text difficulty and predicted text difficulty rises to .85. Such positive confirmation of the predictive accuracy of our regression equation is rather surprising, since strictly speaking the regression equation developed for 5th grade

pupils is only applicable for this grade level). Yet it is obviously possible to predict the empirical difficulty ranking for texts in different kinds of schools and for different grade levels by using the regression equation derived initially for the 5th grade.

One reason for structuring this investigation in the way described was that we were interested in finding out whether C-Tests continued to decrease in difficulty as the pupils got older, and whether this decrease was linear. You may perhaps remember (cf. Raatz & Klein-Bräley 1983) that we had shown such linear decreases in difficulty for the Hauptschule. Figure 2 shows the results of the investigation. For all three types of school the empirical facility values for the C-Tests become successively higher as the age of the pupils increases, and in all three cases the increase is linear.

This has very interesting implications for our regression equations. It is obviously possible to predict the RANK difficulties of texts used in C-Tests with a high degree of accuracy for older pupils even when the regression equation for the 5th grade is used. We have demonstrated that this technique works up to the 10th grade. However, since the regression equation for the 5th grade and that for the 3rd grade make virtually identical predictions, the 5th grade equation can obviously be used for all pupils between the 3rd and the 10th grade. Our data also show, however, that the regression lines for the C-Tests tend to hit the 100% mark at some point near the 11th grade. At the other end of the scale, it is not possible to use C-Tests, which are written tests, much earlier than the end of the 2nd grade in schools. This means that we have a regression equation which predicts the rank difficulty of texts for all relevant age groups of L1 German learners.

However we can go further than this. Two of the tests that were developed for the third grade L1 German learners were administered to our other two groups: to L2 learners and to LF learners. Süßmilch used test D3A in his pilot experiment with the children of foreign workers in Germany (L2 learners). Raatz chose to use test N3 with the LF learners at the Eurozentrum in Cologne.

The results of these tests show a very interesting phenomenon. As Table 2 shows, the rank difficulties of the tests were virtually identical to the rank difficulties derived for the L1 German learners. A further experiment showed that this applied to a group of British Army officers in Mülheim learning German too.

Table 2

Rank difficulties of
texts for various groups

TEXTS	Rank difficulties of texts for various groups			Rank difficulties of texts for various groups		Rank difficulties of texts for various groups
	3rd grade L1	3rd grade L1	3rd grade L1	Adults LF (Cologne)	Adults LF (Army)	3rd grade L2 (Turks)
D3 01	1	1	1	1	1	1
02	2	2	2	2	2	2
03	3	3	3	3	3	3
04	4	4	4	4	4	4
05	5	5	5	5	5	5
TEXTS	3rd grade L1	3rd grade L2 (Turks)	3rd grade L2 (Greeks)	4th grade L2 (Turks)	3rd grade L2 (mixed)	Rank difficulties of texts for various groups
	3rd grade L1	3rd grade L2 (Turks)	3rd grade L2 (Greeks)	4th grade L2 (Turks)	3rd grade L2 (mixed)	
D3A 01	1	1	1	1	1	1
03	2	2	2	2	2	2
20	3	3	3	3	4	4
04	4	4	4	4	3	3
17	6	6	5	5	5	5
05	5	5				

In other words, the regression equation derived for the 5th grade L1 learners may also be used to determine the relative difficulty levels for both L2 and LF groups.

I should like to emphasise the word relative. The absolute difficulty levels can not be derived from our regression equation. However if one or two texts are administered to the relevant groups, the empirically determined difficulties for these texts can be used in conjunction with the regression equation to determine which texts are relatively easier, relatively more difficult, or relatively equal in difficulty with regard to the texts actually examined. This has the added advantage that the texts investigated empirically need not necessarily be included in the test. It will nevertheless be possible to develop tests for specific target groups which will have approximately the right level of difficulty for these groups. While, as I already stated, C-Tests have been shown to function adequately even if they are far too difficult or far too easy, there is no doubt that the best test is one with a medium level of difficulty for the subjects involved. Otherwise factors such as frustration or euphoria play far too large a role.

I think I am probably right in claiming that here for the first time language testers have a technique for estimating test difficulty in advance for the three major groups of test subjects, at least in one language, namely German. I would, however, make one reservation.

Our LF learners were an extremely heterogenous group. They came from a large number of different countries and spoke a large number of mother tongues. In addition their previous learning histories were unknown. The typical group of LF learners, however, is not like this. In the classroom we have a homogeneous monolingual group. In addition this group learns the foreign language - normally at least - through a text book, or teaching materials. Such materials are structured and ordered so as to enable learner to progress through various stages of vocabulary learning, syntax, grammar, communication in a pre-planned sequence. Whether our regression equation will work for groups in a controlled setting has not yet been examined. With L2 groups it seems not unreasonable to assume that they might have met the various components or elements of the language according to their normal frequency of occurrence in a similar fashion to the L1 learners. This assumption cannot be made for learners in a controlled, text-book-determined-language-learning environment. For this reason it may well be the case that the regression equation which is valid for all the other groups might fail to predict difficulty adequately here. The results for the British Army officers seem to contradict this assumption, but these were very advanced LF learners, and at this stage the text book is no longer in control of the language learning process - the advanced learner is generally concerned with authentic language material.

It seems therefore that we have reached our aim in German. What about English?

Once again the largest group of subjects available to us is a German group: in this case LF learners, and most of our research has been concerned with our advanced students of English in Duisburg. This is, incidentally, the group for whom I personally really need techniques for advance determination of difficulty, since my own judgements of texts involved almost invariably turn out to be wrong. I nearly always overestimate the difficulty, and the test turns out to be too easy in the empirical trials.

Once again we followed the same procedure: empirical test trials, calculation of the various indices, determination of the regression equation. We decided to use the C-Test values for the first group for whom the test had turned out to have an appropriate level of difficulty. These were the 67 students who were tested on entering the university in October 1982.

The regression equation for English texts for LF learners, calculated on the basis of 6 texts, once again uses the type-token ratio, but the syntax predictor here is the average sentence length in syllables. The equation takes the form:

$$Y' = 192 - 133 \text{ TTR} - 0.7 \text{ XSL}(S)$$

Table 3
Prediction for LF groups

Text	P' values	P values DU 82	P values DU 83	P values SF
GB07	97	94		88
GB11	66	66	75	55
GB17	80	83		67
GB20	60	58	62	50
GB19	63	59	64	44
GB18	56	56		55
GB30	81		81	
$r_{12} = .987$ $r_{13} = 1.00$ $r_{14} = .930$				

Here the beta weights are -.65 for the TTR and -.83 for the mean sentence length in syllables. The multiple correlation between these two indices and the empirically determined text difficulties is a staggering .98!

Again, however, we need to determine the accuracy of the predictions obtained. One proof lies in the data obtained by using some of the same texts in the tests run for a new group on October 4th 1983. For the text GB 20 the predicted value is 60 and the empirically measured value is 62. For text GB 19 the predicted value is 63 and the empirical value 64. For text GB 11 the predicted value is 66 and the empirical value 75. The fourth text in this test (GB 30) had never been used before, but had a predicted value of 81. Its empirical value turned out to be precisely 81 (cf. Table 3). The rank correlation coefficient between the predicted difficulty levels and the empirically determined levels is 1.00.

This seems to indicate that the regression equation calculated for incoming students in Duisburg works very well for this group. I was fortunately able to make a second check. Around three weeks ago I received a set of tests from Hilka von Essen which she had admin-

Figure 3 Predicted and empirical difficulty values for Finnish LF group

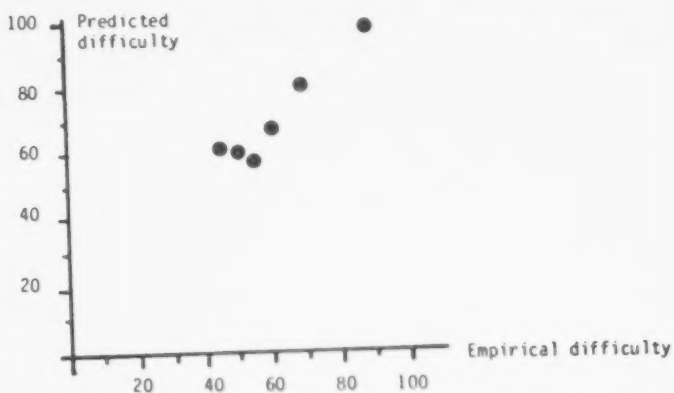
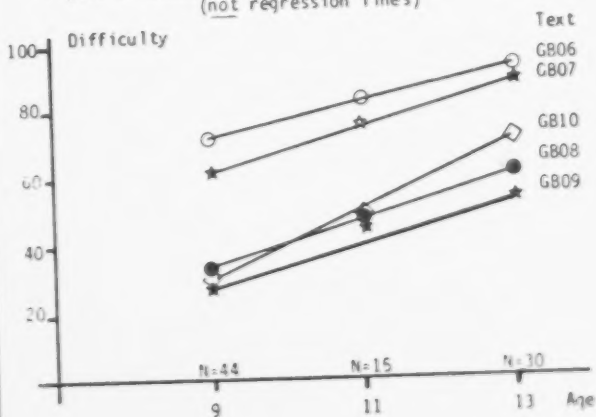


Figure 4 Linearity of difficulty for English L1 groups
(not regression lines)



istered to her students in Finland (SF group in Table 3). This was the same test which had originally been used to develop the regression equation. Here therefore we can check on the accuracy of the prediction for a different LF advanced group. The correlation between the empirical values and the predicted values in this case is .93, and as Figure 3 shows, again we have a case where obviously what is needed is simply an additive constant.

A final confirmation is given by the use of the regression equation for LF groups to predict text difficulty for L1 groups. Here three factors have changed: the group for whom the prediction is to be used is L1 and not LF; the age levels are different (9 to 13 year olds and not 20 year olds) and the texts are entirely different. Again, it is possible to show that the decrease in difficulty is linear (Figure 4). In all cases the correlation between the predicted difficulties and the actual difficulties is between .6 and .7. With only 5 texts involved this value is, of course, not statistically significant. But as examination of the data shows, (Figure 5), the prediction obtained from the regression equation successfully divide the texts into two groups, and they do this for all three age levels involved. It would therefore be possible to work with this equation until we have more stable data from larger L1 groups.

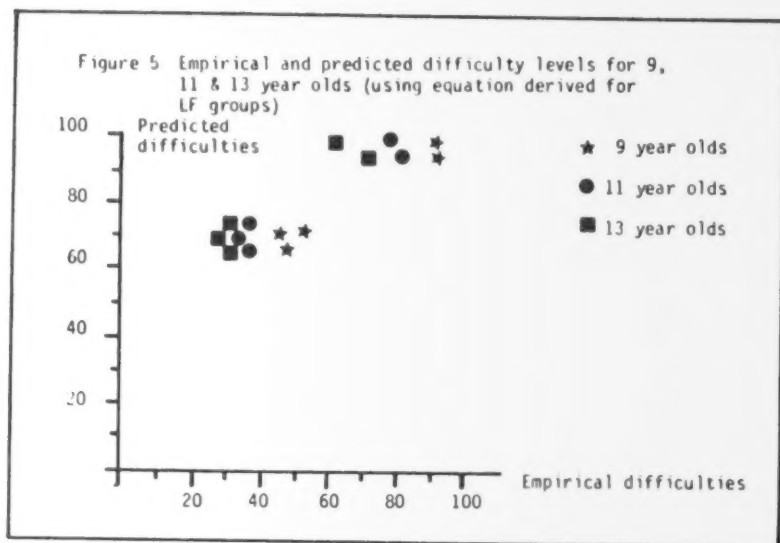
The question I have started to ask myself as a result of these findings for both English and German is whether the basic equation ought perhaps to be the one for the L1 groups. But more research will be necessary before this question can be answered.

What conclusions can be drawn from all this? The first is an eminently practical one, namely that one of the major problems with the C-Test, that of advance prediction of text difficulty, is well on the way to being solved. In our future work we shall be making use of these equations, and at a future IUS meeting we shall report on their performance.

The second conclusion is more theoretical. It seems to me that these results imply that there is such a thing as absolute text difficulty, which I propose to call C-processing difficulty. We have shown both for English and German that the C-processing difficulty decreases in a linear fashion as the L1 subjects get older. We have shown that regression equations using text statistics derived for specific subgroups can be used with a fair degree of predictive accuracy over wide age ranges. We have also shown that the same regression equations can be used for L1 and LF groups, and that for German they can also be used for L2 groups.

One of the things that fascinate me about our results is the parallelism for the different learner groups. This seems to me to have all sorts of implications for language acquisition research. However I have, as yet, only begun to scratch the surface of this problem. Instead I should like to finish by reporting our newest idea.

You will remember that the advantages of the Rasch model were



firstly that the latent trait of the subject and the difficulty of the items can be measured unidimensionally on an interval scale and secondly that the latent trait can be measured independently of the items and vice versa. In the Rasch model the parameters ξ and ϵ are related to each other in a complex multiplicative function (logistic function). For quite a while now we have been searching for a way to use the Rasch model for our data. The observed score for any individual on a C-Test must be dependent on the one hand on the C-processing difficulty of the text, and on the other on the individual's position on the language learning continuum. A simpler formal model than the Rasch model for this relationship could be an additive one:

$$\text{observed score} = \text{subject ability} + \text{text level} + \text{error}.$$

In October 1983 at a psychology congress in Düsseldorf Moosbrugger and Müller presented their CLA model, which has precisely this basic equation. The model is analogous to the Rasch model; it has exactly the same qualities and is intended to be used for at least ordinally scaled data.

Here the parameters are $X_{ij} = \xi_i + \lambda_j + \epsilon_{ij}$ where ξ_i is subject ability (latent trait) and λ_j is text difficulty level (C-processing difficulty).

We intend to use this model and hope that it will enable us to disentangle C-processing difficulty and subject ability, positing a latent trait which we might call general language competence. If this is possible, then the C-Test which began its life simply as a toy for two incurable test freaks may well turn out to represent a significant advance in research into language learning.

Notes

I should like to thank Werner Debertin who administered the tests as part of his thesis for the First State Examination for Teachers.

Bibliography

- Alderson, J. C. 1979. "The cloze procedure and proficiency in English as a foreign language". TESOL Quarterly 13,2: 219-226.
- Bartenwerfer, H. and U. Raatz. 1979. Einführung in die Psychologie 6. Methoden der Psychologie. Wiesbaden: Akademische Verlagsgesellschaft.
- Bensoussan, M. 1982. "Testing the test of advanced EFL reading comprehension. To what extent does the difficulty of a multiple-choice reading comprehension test reflect the difficulty of the text?" SYSTEM, 10: 285-290.
- 1983. Changing the difficulty level of multiple-choice reading comprehension questions. Paper presented at the ACROLY Symposium, Qufriyat Anavim, Israel.

- Herdan, G. 1966. The advanced theory of language as choice and chance. Berlin: Springer.
- Klein-Braley, C. 1981a. Empirical investigations of cloze tests. Dissertation, Duisburg.
- 1981b. The determination of text difficulty. Paper presented at the 12th Jahrestagung of the Gesellschaft für Angewandte Linguistik, Mainz.
- Moosbrugger, H. 1983. Neuere psychologische Testtheorie und ihre Anwendung in der Psychodiagnostik. Paper presented to the 12th Kongreß für Angewandte Psychologie des BDP, Düsseldorf.
- Oller, J.W.Jr. 1979. Language tests at school. London: Longman.
- Pugh, A.K. 1981. "Text analysis, comprehension and reading behaviour". In: Hopkins, E. and R. Grotjahn. (eds). 1981. Studies in language teaching and language acquisition. Quantitative Linguistics 9. Bochum: Brockmeyer. 109-128.
- Stokes, A. 1980. The reliability of readability formulae. In: Pugh, A.K., Lee, V.J. and J. Swann. (eds). 1980. Language and language use. London: Heinemann. 345-357.